REMARKS

The Examiner has rejected claims 1-31. Claims 1-31 are pending for examination with claims 1,8,18, and 22 being independent claims.

The Examiner has rejected Claim 1 and 18 under 35 U.S.C. §103(a) as being unpatentable over U.S. Patent No. WO 99/62007 to Fayyad et al. ("Fayyad") in view of U.S. Patent No. 6,470,344 to Kothuri et al. ("Kothuri").

Applicants have amended Claim 1 to call for:

> "performing clustering of data records in two phases including a first phase and a second phase, the first phase clustering the data records over a discrete attribute space using an itemset identification, and the second phase clustering continuous attributes using a method for clustering continuous attribute data the data to produce an intermediate set of data clusters" (underlining added for emphasis)

Applicants have amended Claim 8 to call for:

> "performing a first discrete cluster by counting data records from the database which have the same discrete attribute configuration itemset identification" (underlining added for emphasis)

Applicants have amended Claim 18 to call for:

> "a first clustering of data records having the same or a similar specified discrete attribute configuration performed by itemset identification" (underlining added for emphasis)

Applicants have amended Claim 22 to call for:

> "performing a first clustering of data records from the database which have specified discrete attribute configurations by itemset identification" (underlining added for emphasis)

As such, Applicants submit that Claim 1, 8, 18 and 22 are not unpatentable over Fayyad in view of Kothuri.

The present invention provides:

> "The data mining engine 12 clusters data records stored on a database 10 made up of data records having multiple attributes or fields that contain both discrete and continuous data." (page 5, line 30 to page 6, line 1) (underlining added for emphasis).
>
> and
>
> "In accordance with the present invention, the clustering model 15 is arrived at in two phases. A cluster structure over the discrete attribute space is first performed using methods similar to methods for identifying frequent itemsets in data. Known frequent itemset identification algorithms are efficient in dealing with 1000 - 100,000s of attributes. The present invention uses similar methods to locate discrete attribute cluster structure. Once this cluster structure is determined, structure over the continuous attributes of data is identified using one of the many methods currently available for clustering continuous attribute data. " (page 8:lines 1-8) (underlining added for emphasis).

Kothuri, on the other hand provides for:

> " In one effective application of a VAMSplit algorithm, a set of N data points is divided whenever N>M (i.e., there are too many data points to fit in a single node). In order to divide the dataset most equitably, the distribution of data values within each dimension is computed (e.g., the difference between the smallest and greatest values is determined). The data is then sorted in the dimension having the greatest variance and the sorted dataset is then divided in that dimension as close to a median value as possible." (col. 11, lines 33-41), (underlining added for emphasis)
>
> and

“ After the first partition, it is determined whether each new subset of data will fit into a single R-tree node. Each subset that is too large (i.e., each subset that contains more than M data points) is further divided in a similar manner.

Thus, a first data subset (e.g., that which includes values in the x dimension less than m(11)) is further subdivided as follows. This data subset possesses greater variance in the y dimension than the x dimension, and the approximate median (denoted by m(6)) for the subset is computed accordingly. Because N=2*M,

$$m(6)=3*\text{floor}(6/(2*)+0.5)=3$$

and the first subset is divided after the third data point in the y dimension. Dividing line 312 illustrates this division.

The second subset (i.e., the subset having values in the x dimension greater than m.sub.1) yields an approximate median value in the x dimension of 2 (represented on the x-axis by m(5)). In particular, N<2*M, therefore

$$m(5)=\text{floor}(5/2)=2$$

and the second subset is divided after the second data point. This division is represented by dividing line 314. It will be apparent that each of the approximate median values computed above is relative to the subject dataset or data subset. Thus, the representations of m(11), m(6) and m(5) on the x- and y-axes should be considered relative to the origin or the preceding approximate median value, as appropriate.

After the second subset is divided, each subdivision, or cluster, of data points can now fit into a node of an R-tree having a fanout value of three; therefore no further division is required. Clusters 320, 322, 324, 326 in FIG. 3 are demarcated by dotted boundaries. Now that the data is clustered appropriately, the R-tree index may be constructed by placing each cluster of data items into a separate leaf node (e.g., by placing identifiers of each data item and its storage location or address in the appropriate leaf node). From the leaf nodes, parent nodes may be formed such that each entry in a parent node comprises an MBA of a cluster of data and an identifier of (e.g., a pointer to) the corresponding leaf node. In similar fashion, grandparent nodes of the leaf nodes, and higher nodes as necessary, may be formed. Eventually a root node of the R-tree is constructed." (col. 12, lines 12-54), (underlining added for emphasis)

and

"An illustrative algorithm for clustering multi-attribute data items for leaf nodes of an R-tree index is now provided: 1. Store all data items in a single node if possible (i.e., depending upon the fanout of the index) 2. Otherwise, until each subset or cluster of data items will fit into a single node, do: 2.1 Select a dimension in which to divide an over-populated cluster or subset of data items: 2.1.1 Compute the number of query retrieval units in each dimension: 2.1.1.1 For attributes having discrete values, determine the number of possible values (e.g., number of cities for a region dimension or brand names for a product dimension) 2.1.1.2 For attributes having a continuous range of values (e.g., latitude), a query pattern may specify a percentage, P, of the values in that domain that may be accessed in a particular query; the number of query retrieval units is then equal to (100/P)*(range of subset/range of entire set) 2.1.1.3 Some attributes (e.g., time) may be represented by discrete values (e.g., days, months, quarters) or a continuous range; the number of query retrieval units is measured accordingly 2.1.2 The attribute that has the most query retrieval units is selected as the dimension in which to divide the data items 2.2 Sort the data items in the selected dimension 2.3 Divide the data items, possibly to yield an (approximately) equal number of data items in each subset or, as one alternative, in a manner that yields a number of data subsets one or more of which will fit into individual leaf nodes 3. Repeat

> steps 1 and 2 until each subset fits into a single node."
> (col. 14, lines 30-65), (underlining added for emphasis)

Accordingly, Applicants submit that Claim 1, 8, 18 and 22 are not unpatentable over Fayyad in view of Kothuri.

Claims 2-7 are dependent on Claim 1. As such, Claims 2-7 are believed allowable based upon Claim 1.

Claims 9-17 are dependent on Claim 8. As such, Claims 9-17 are believed allowable based upon Claim 8.

Claims 19-21 are dependent on Claim 18. As such, Claims 19-21 are believed allowable based upon Claim 18.

Claims 23-31 are dependent on Claim 22. As such, Claims 23-31 are believed allowable based upon Claim 22.

CONCLUSION

Accordingly, in view of the above amendment and remarks it is submitted that the claims are patentably distinct over the prior art and that all the rejections to the claims have been overcome. Reconsideration and reexamination of the above Application is requested. Based on the foregoing, Applicants respectfully requests that the pending claims be allowed, and that a timely Notice of Allowance be issued in this case. If the Examiner believes, after this amendment, that the application is not in condition for allowance, the Examiner is requested to call the Applicant's attorney at the telephone number listed below.

If this response is not considered timely filed and if a request for an extension of time is otherwise absent, Applicants hereby request any necessary extension of time. If there is a fee occasioned by this response, including an extension fee that is not covered by an enclosed check please charge any deficiency to Deposit Account No. 50-0463.

Respectfully submitted,

Microsoft Corporation

Date: 5/15/06

By: ______

Microsoft Corporation
One Microsoft Way
Redmond, WA 98052-6399

Paul B. Heynssens, Reg. No.: 47,648
Attorney for Applicants
Direct telephone (425) 707-3913

CERTIFICATE OF MAILING OR TRANSMISSION
(Under 37 CFR § 1.8(a)) or ELECTRONIC FILING

I hereby certify that this correspondence is being electronically deposited with the USPTO via EFS-Web on the date shown below:

May 15, 2006
Date

Signature

Noemi Tovar
Printed Name